

# Spectral Shape of Check-Hybrid GLDPC Codes

Enrico Paolini, Mark F. Flanagan, Marco Chiani and Marc P. C. Fossorier

**Abstract**—This paper analyzes the asymptotic exponent of both the weight spectrum and the stopping set size spectrum for a class of generalized low-density parity-check (GLDPC) codes. Specifically, all variable nodes (VNs) are assumed to have the same degree (regular VN set), while the check node (CN) set is assumed to be composed of a mixture of different linear block codes (hybrid CN set). A simple expression for the exponent (which is also referred to as the *growth rate* or the *spectral shape*) is developed. This expression is consistent with previous results, including the case where the normalized weight or stopping set size tends to zero. Furthermore, it is shown how certain symmetry properties of the local weight distribution at the CNs induce a symmetry in the overall weight spectral shape function.

## I. INTRODUCTION

Tanner codes were introduced in [1] as a generalization of Gallager's low-density parity-check (LDPC) codes [2]. In the bipartite graph representation of a Tanner code, all variable nodes (VNs) have the same degree and may be interpreted as repetition codes with the same length (regular VN set). Moreover, all check nodes (CNs) are generic linear block codes with the same length, dimension and code book (regular CN set). Tanner codes with an *irregular* VN set (i.e., VNs with different degrees) or with a *hybrid* CN set (i.e., CNs of different types) are more generically referred to as generalized LDPC (GLDPC) codes in the literature (e.g. [3]). Note that an LDPC code may be viewed as a GLDPC code where all CNs are single parity-check (SPC) codes. An even more general class of codes is represented by doubly-generalized LDPC (D-GLDPC) codes [4], where the VNs are also allowed to be generic linear block codes.

In this paper, a simple formula for the asymptotic (in codeword length) exponent of the weight distribution of a GLDPC code ensemble with a regular VN set and a hybrid CN set is developed. As usual in the literature, this exponent will be referred to as the *growth rate of the weight distribution* or the *weight spectral shape* of the ensemble, the two expressions being used interchangeably throughout the paper. The starting point for deriving the above-mentioned formula is a polynomial system solution for the spectral shape that was developed by the authors in [5, Theorem 1]. Here, it was shown that any value of the spectral shape function of an irregular D-GLDPC code ensemble can be calculated by solving a  $(4 \times 4)$  system of polynomial equations, regardless of the number of different

VN and CN types. Additional recent works relevant to the subject of this paper are [6]–[10].

As explained in Section II, assuming transmission over the binary erasure channel (BEC) and iterative decoding, the developed formula is also valid for the asymptotic exponent of the stopping set size distribution upon replacing the local weight enumerating function (WEF) of each CN type with an appropriate polynomial function.

Symmetry properties of the growth rate of the weight distribution are also investigated. It is proved that the weight spectral shape function of a variable-regular GLDPC ensemble is symmetric w.r.t. normalized weight  $\alpha = 1/2$  if the local WEF of each CN is a symmetric polynomial. This result establishes a connection between symmetry properties at a “microscopic” level (i.e., at the nodes of the Tanner graph) and symmetry of the “macroscopic” growth rate function. A necessary condition for symmetry of the weight spectral shape is also developed.

## II. PRELIMINARY DEFINITIONS

We consider a check-hybrid GLDPC code ensemble  $\mathcal{M}_n$ , where  $n$  is the codeword length (this is equal to the number of VNs). All VNs are repetition codes of length  $q \geq 2$ , with input-output weight enumerating function

$$B(x, y) = 1 + xy^q. \quad (1)$$

There are  $n_c$  different CN types  $t \in I_c = \{1, 2, \dots, n_c\}$ ; for each CN type  $t \in I_c$ , we denote by  $h_t$ ,  $s_t$  and  $r_t$  the CN dimension, length and minimum distance, respectively. We assume that  $r_t \geq 2$  for all  $t \in I_c$ , and that no CN has idle bits (i.e., its generator matrix contains no all-zero column). The WEF for CN type  $t \in I_c$  is given by

$$A^{(t)}(z) = \sum_{u=0}^{s_t} A_u^{(t)} z^u = 1 + \sum_{u=r_t}^{s_t} A_u^{(t)} z^u.$$

Here  $A_u^{(t)} \geq 0$  denotes the number of weight- $u$  codewords for CNs of type  $t$ . We denote by  $\bar{u}_t$  the largest  $u \in \{r_t, r_t + 1, \dots, s_t\}$  such that  $A_u^{(t)} > 0$ .

For  $t \in I_c$ ,  $\rho_t$  is the fraction of edges of the Tanner graph connected to type- $t$  CNs, and the polynomial  $\rho(x)$  is defined by

$$\rho(x) = \sum_{t \in I_c} \rho_t x^{s_t-1}.$$

If  $E$  denotes the number of edges in the Tanner graph, the number of CNs of type  $t \in I_c$  is then given by  $E\rho_t/s_t$ . Denoting as usual  $\int_0^1 \rho(x) dx$  by  $\int \rho$ , we see that the number of edges in the Tanner graph is given by  $E = nq$  and the number

E. Paolini and M. Chiani are with DEIS/WiLAB, University of Bologna, via Venezia 52, 47023 Cesena (FC), Cesena, Italy (e-mail: e.paolini@unibo.it, marco.chiani@unibo.it).

M. F. Flanagan is with the School of Electrical, Electronic and Mechanical Engineering, University College Dublin, Belfield, Dublin 4, Ireland (e-mail: mark.flanagan@ieee.org).

M. P. C. Fossorier is with ETIS ENSEA, UCP, CNRS UMR-8051, 6 avenue du Ponceau, 95014 Cergy Pontoise, France (e-mail: mfossorier@ieee.org).

of CNs is given by  $m = E \int \rho$ . Therefore, the fraction of CNs of type  $t \in I_c$  is given by

$$\gamma_t = \frac{\rho_t}{s_t \int \rho}. \quad (2)$$

A member of the GLDPC code ensemble  $\mathcal{M}_n$  corresponds to a permutation of the  $E$  edges connecting VNs to CNs.

In the special case where there is only *one* CN type (Tanner code ensemble), we write the WEF for this CN type as  $A(z) = \sum_{u=0}^s A_u z^u = 1 + \sum_{u=r}^s A_u z^u$ . In this case, the largest  $u \in \{r, r+1, \dots, s\}$  such that  $A_u > 0$  is denoted by  $\bar{u}$ .

The weight spectral shape of a GLDPC code ensemble sequence  $\{\mathcal{M}_n\}$  is defined by

$$G(\alpha) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_{\mathcal{M}_n} [N_{\alpha n}] \quad (3)$$

where  $\mathbb{E}_{\mathcal{M}_n}$  denotes the expectation operator over the ensemble  $\mathcal{M}_n$ ,  $N_w$  denotes the number of codewords of weight  $w$  of a randomly chosen GLDPC code in the ensemble and the logarithm has base  $e$ . The limit in (3) assumes the inclusion of only those positive integers  $n$  for which  $\alpha n \in \mathbb{Z}$  and  $\mathbb{E}_{\mathcal{M}_n} [N_{\alpha n}]$  is positive. Using standard notation, we also define the asymptotic relative minimum distance for  $\mathcal{M}_n$  as  $\alpha^* = \inf\{\alpha > 0 \mid G(\alpha) \geq 0\}$ . The ensemble sequence is said to exhibit *good spectral shape behavior* when  $\alpha^* > 0$  and *bad spectral shape behavior* when  $\alpha^* = 0$ .

Although this paper is focused on the weight spectrum, the results developed in Section III can be extended to the stopping set size spectrum. A stopping set of a GLDPC code may be defined as any subset  $\mathcal{S}$  of the VNs such that, assuming all VNs in  $\mathcal{S}$  are erased and all VNs not in  $\mathcal{S}$  are not erased, every CN which is connected to  $\mathcal{S}$  cannot recover any VN in  $\mathcal{S}$ .<sup>1</sup> A *local stopping set* for a CN is a subset of the local code bits which, if erased, is not recoverable to any extent by the CN. All results derived in this paper for the distance spectrum can be extended to the stopping set size spectrum by simply replacing the WEF for CN type  $t \in I_c$  with its local stopping set enumerating function (SSEF).

We point out that the local SSEF of a CN depends on the decoding algorithm used to locally recover from erasures. In this paper, we will consider both bounded distance (BD) and maximum *a posteriori* (MAP) CN decoding. In the former case, the local SSEF (BD-SSEF) is given by

$$\Psi^{(t)}(z) = 1 + \sum_{u=r_t}^{s_t} \binom{s_t}{u} z^u. \quad (4)$$

In the latter case, the local SSEF (MAP-SSEF) is given by

$$\Phi^{(t)}(z) = 1 + \sum_{u=r_t}^{s_t} \phi_u^{(t)} z^u \quad (5)$$

where  $\phi_u^{(t)} \geq 0$  is the number of local stopping sets (under MAP decoding) of size  $u$ .<sup>2</sup>

<sup>1</sup>The concept of stopping set was first introduced in [11] in the context of LDPC codes. When applied to LDPC codes (i.e., all CNs are SPC codes), the definition of stopping set used in this paper coincides with that in [11].

<sup>2</sup>Denoting by  $\mathbf{G}_t$  any generator matrix for a type- $t$  CN, a local erasure pattern is a local stopping set under MAP decoding when each column of  $\mathbf{G}_t$  corresponding to erased bits is linearly independent of the columns of  $\mathbf{G}_t$  corresponding to the non-erased bits.

The growth rate of the stopping set size distribution of the ensemble sequence  $\{\mathcal{M}_n\}$  for the case of BD and MAP decoding at the CNs, whose definition is analogous to (3), will be denoted by  $G_\Psi(\alpha)$  and  $G_\Phi(\alpha)$ , respectively. Similarly the asymptotic relative minimum stopping set size will be denoted by  $\alpha_\Psi^*$  and  $\alpha_\Phi^*$ , respectively.

*Definition 1:* Let

$$M \triangleq \left( \int \rho \right) \sum_{t \in I_c} \gamma_t \bar{u}_t \leq 1 \quad (6)$$

and define the function  $f : \mathbb{R}^+ \rightarrow [0, M)$  as

$$f(z) = \left( \int \rho \right) \sum_{t \in I_c} \gamma_t \frac{z \frac{dA^{(t)}(z)}{dz}}{A^{(t)}(z)}. \quad (7)$$

Note that we have  $M = 1$  if and only if  $\bar{u}_t = s_t$  for all  $t \in I_c$ .

*Lemma 1:* The function  $f$  fulfills the following properties:

- 1) It is monotonically increasing for all  $z > 0$ ;
- 2)  $f(0) = f'(0) = 0$ ;
- 3)  $\lim_{z \rightarrow +\infty} f(z) = M$ .

*Proof:* We prove the first property, as the proofs of the second and the third properties are straightforward. The derivative of  $f$  (normalized w.r.t.  $\int \rho$ ) is given by

$$\sum_{t \in I_c} \gamma_t \frac{A^{(t)}(z) \left[ \frac{dA^{(t)}(z)}{dz} + z \frac{d^2 A^{(t)}(z)}{dz^2} \right] - z \left[ \frac{dA^{(t)}(z)}{dz} \right]^2}{[A^{(t)}(z)]^2}.$$

The denominator of the fraction in each term in the sum is strictly positive for all  $z > 0$ . The numerator of the fraction in term  $t \in I_c$  in the sum may be expanded as

$$\begin{aligned} & \left( 1 + \sum_{v=r_t}^{s_t} A_v^{(t)} z^v \right) \left( \sum_{u=r_t}^{s_t} u A_u^{(t)} z^{u-1} + \sum_{u=r_t}^{s_t} u(u-1) A_u^{(t)} z^{u-1} \right) \\ & - z \left( \sum_{u=r_t}^{s_t} u A_u^{(t)} z^{u-1} \right) \left( \sum_{v=r_t}^{s_t} v A_v^{(t)} z^{v-1} \right) \\ & = \sum_{u=r_t}^{s_t} u^2 A_u^{(t)} z^{u-1} + \sum_{u=r_t}^{s_t} \sum_{v=r_t}^{s_t} u(u-v) A_u^{(t)} A_v^{(t)} z^{u+v-1}. \end{aligned}$$

Observe that in this expression, each term in the second summation with  $u = v$  is zero, while each  $(u, v)$  term in the second summation (with  $u > v$ ) added to the corresponding  $(v, u)$  term is positive for  $z > 0$ , since  $u(u-v) A_u^{(t)} A_v^{(t)} z^{u+v-1} + v(v-u) A_v^{(t)} A_u^{(t)} z^{u+v-1} = (u-v)^2 A_u^{(t)} A_v^{(t)} z^{u+v-1} > 0$  and therefore the second summation is nonnegative for  $z > 0$ . Since the first summation is strictly positive for  $z > 0$ , it follows that  $f'(z) > 0$  for all  $z > 0$ .  $\square$

Note that, due to Lemma 1, the inverse of  $f$ , denoted by  $f^{-1} : [0, M) \rightarrow \mathbb{R}^+$ , is well-defined.

### III. SPECTRAL SHAPE OF CHECK-HYBRID GLDPC CODES

We next state and prove an expression for the spectral shape of check-hybrid GLDPC codes.

*Theorem 1 (Spectral shape of check-hybrid GLDPC codes):* Consider a GLDPC code ensemble with a regular VN set, composed of repetition codes all of length  $q$ , and a hybrid CN set, composed of a mixture of  $n_c$  different linear block

code types. Then, the weight spectral shape of the ensemble is given by

$$G(\alpha) = (1 - q)h(\alpha) - q\alpha \log f^{-1}(\alpha) + q \left( \int \rho \right) \sum_{t \in I_c} \gamma_t \log A^{(t)}(f^{-1}(\alpha)) \quad (8)$$

where  $h(\alpha) = -\alpha \log \alpha - (1 - \alpha) \log(1 - \alpha)$  denotes the binary entropy function.

*Proof:* In [5], a polynomial system solution for the spectral shape of irregular D-GLDPC codes was derived. In the special case of D-GLDPC codes where all VNs are repetition codes of length  $q$ , this is given by (special case of equation (8) in [5])

$$G(\alpha) = \log B(x_0, y_0) - \alpha \log x_0 + q \left( \int \rho \right) \sum_{t \in I_c} \gamma_t \log A^{(t)}(z_0) + q \log(1 - \frac{\beta}{q}) \quad (9)$$

where the values of  $x_0, y_0, z_0, \beta$  in (9) are found by solving the  $(4 \times 4)$  polynomial system

$$\left( \int \rho \right) \sum_{t \in I_c} \gamma_t \frac{z_0 \frac{dA^{(t)}(z_0)}{dz}}{A^{(t)}(z_0)} = \frac{\beta}{q}, \quad (10)$$

$$\frac{x_0 y_0^q}{1 + x_0 y_0^q} = \alpha, \quad (11)$$

$$\frac{x_0 y_0^q}{1 + x_0 y_0^q} = \frac{\beta}{q}, \quad (12)$$

and

$$\frac{z_0 y_0}{1 + z_0 y_0} = \frac{\beta}{q}. \quad (13)$$

Note that we are certain of the existence of a unique real solution to the polynomial system such that  $x_0 > 0, y_0 > 0, z_0 > 0, \beta > 0$ , due to Hayman's formula. We solve this system of equations sequentially for the variables  $\beta, z_0, y_0$  and  $x_0$  (respectively). First, combining (11) and (12) yields

$$\beta = q\alpha. \quad (14)$$

Substituting (14) into (10) yields  $f(z_0) = \alpha$  which may be written as

$$z_0 = f^{-1}(\alpha). \quad (15)$$

Using (14) and (15) in (13) yields

$$y_0 = \frac{\alpha}{(1 - \alpha)f^{-1}(\alpha)}. \quad (16)$$

Finally, substituting (14) and (16) into (12) yields

$$x_0 = \left( \frac{\alpha}{1 - \alpha} \right)^{1-q} (f^{-1}(\alpha))^q. \quad (17)$$

Substituting (14), (15), (16) and (17) into (9), and simplifying, leads to (8).  $\square$

The expression (8) holds regardless of whether the ensemble has good or bad spectral shape behavior. Note that, according to (8), the growth rate  $G(\alpha)$  is well-defined only for  $\alpha \in [0, M]$ . This is as expected due to the following reasoning. A codeword of weight  $\alpha n$  naturally induces a distribution of bits on the Tanner graph edges,  $\alpha n q$  of which are equal to 1. Also

note that the maximum number of ones in this distribution occurs when a maximum weight local codeword is activated for each of the  $\gamma_t m$  CNs of type  $t \in I_c$ , and is thus given by  $m \sum_{t \in I_c} \gamma_t \bar{u}_t$ . Hence, we have  $\alpha n q \leq m \sum_{t \in I_c} \gamma_t \bar{u}_t$ , i.e.,  $\alpha \leq M$ .

In Appendix I it is shown how, for small relative weight  $\alpha$ , (8) simplifies to a known expression that was derived in [12] for Tanner codes, and extended in [13] to irregular GLDPC codes.

By considering Theorem 1 in the special case of Tanner codes, we obtain the following corollary.

*Corollary 1 (Spectral shape of Tanner codes):* Consider a Tanner code ensemble where all variable component codes are length- $q$  repetition codes and where all check component codes are length- $s$  codes with weight enumerating function  $A(z) = 1 + \sum_{u=r}^s A_u z^u$ . The weight spectral shape of this ensemble is given by

$$G(\alpha) = (1 - q)h(\alpha) - q\alpha \log(f^{-1}(\alpha)) + \frac{q}{s} \log A(f^{-1}(\alpha)) \quad (18)$$

where the function  $f$  is given by (special case of (7))

$$f(z) = \frac{z A'(z)}{s A(z)}, \quad (19)$$

and  $f^{-1} : [0, M] \rightarrow \mathbb{R}^+$  is well-defined, where  $M = \frac{\bar{u}}{s}$ .

Note that, in the special case where all CNs are SPC codes, (18) becomes equal to the spectral shape expression for regular LDPC codes developed in [16, Theorem 2] for the case of stopping sets. Also note that, in some cases, (18) can be expressed analytically as  $f^{-1}(\alpha)$  admits an analytical form. An example is given in Appendix II.

#### IV. SYMMETRY OF THE WEIGHT SPECTRAL SHAPE

Consider a GLDPC code ensemble with a regular VN set and a hybrid CN set. In this section, we show how a symmetry in the overall weight spectral shape of the ensemble is induced by local symmetry properties in the WEFs of the CNs.

*Definition 2:* For CN type  $t \in I_c$ , let  $U_t = \{u \in \mathbb{N} | A_u^{(t)} > 0\}$ . Then, we define

$$U^{(t*)} = \{v \in \mathbb{N} | \bar{u}_t - v \in U_t\}. \quad (20)$$

Note that for any CN type, we always have  $0 \in U^{(t*)}$  and  $\bar{u}_t \in U^{(t*)}$ .

*Definition 3:* The WEF of CN type  $t \in I_c$  is said to be *symmetric* if and only if  $A_{\bar{u}_t - u}^{(t)} = A_u^{(t)}$  for all  $u \in U_t$ .

Note that if the WEF of CN type  $t \in I_c$  is symmetric, then we have  $U^{(t*)} = U^{(t)}$ .

*Lemma 2:* The WEF of CN type  $t \in I_c$  is symmetric if and only if the all-1 codeword belongs to this code.

The proof of Lemma 2 is omitted due to space constraints.<sup>3</sup>

<sup>3</sup>The sufficient condition (if a linear block code has the all-1 codeword then its WEF is symmetric) is a well-known result in classical coding theory. On the other hand, we proved the necessary condition by assuming that CN type  $t \in I_c$  has a symmetric WEF and a maximum codeword weight  $D < s_t$ , and by showing that these assumptions lead to a contradiction. As pointed out by one of the anonymous reviewers, the necessary condition may also be proved by reasoning on the expected weight of a randomly selected codeword in a linear block code, under the symmetry hypothesis.

*Lemma 3:* The WEF of CN type  $t \in I_c$  fulfills

$$A^{(t)}(z) = z^{\bar{u}_t} A^{(t)}(z^{-1}), \quad (21)$$

for all  $z \in \mathbb{R}^+$  if and only if it is symmetric.

*Proof:* We have

$$A^{(t)}(z^{-1}) = \frac{\sum_{u \in U_t} A_u^{(t)} z^{\bar{u}_t - u}}{z^{\bar{u}_t}} = \frac{\sum_{v \in U^{(t*)}} A_{\bar{u}_t - v}^{(t)} z^v}{z^{\bar{u}_t}}$$

where the final equality is obtained by  $v = \bar{u}_t - u$ . The proof is completed by observing that, if and only if  $A^{(t)}(z)$  is symmetric, we have  $U^{(t*)} = U_t$  and  $A_{\bar{u}_t - v}^{(t)} = A_v^{(t)}$  for all  $v \in U_t$ .  $\square$

*Lemma 4:* The function  $f$  defined by (7) fulfills

$$f(z) = M - f(z^{-1}) \quad (22)$$

$\forall z \in \mathbb{R}^+$  if and only if  $A^{(t)}(z)$  is symmetric for every  $t \in I_c$ .

*Proof:* First we note that if and only if the WEF of CN type  $t \in I_c$  is symmetric, we have, differentiating (21),

$$\frac{dA^{(t)}(z)}{dz} = -z^{\bar{u}_t - 2} \frac{dA^{(t)}(z^{-1})}{dz^{-1}} + \bar{u}_t z^{\bar{u}_t - 1} A^{(t)}(z^{-1}).$$

Multiplying by  $z$  and using (21) yields

$$z \frac{dA^{(t)}(z)}{dz} = -z^{\bar{u}_t - 1} \frac{dA^{(t)}(z^{-1})}{dz^{-1}} + \bar{u}_t A^{(t)}(z). \quad (23)$$

Then,

$$\begin{aligned} M - f(z^{-1}) &= M - \left( \int \rho \right) \sum_{t \in I_c} \gamma_t \left( \frac{z^{-1} \frac{dA^{(t)}(z^{-1})}{dz^{-1}}}{A^{(t)}(z^{-1})} \right) \\ &\stackrel{(a)}{=} \left( \int \rho \right) \sum_{t \in I_c} \gamma_t \left( \bar{u}_t - \frac{z^{\bar{u}_t - 1} \frac{dA^{(t)}(z^{-1})}{dz^{-1}}}{A^{(t)}(z)} \right) \\ &\stackrel{(b)}{=} \left( \int \rho \right) \sum_{t \in I_c} \gamma_t \frac{\bar{u}_t A^{(t)}(z) + z \frac{dA^{(t)}(z)}{dz} - \bar{u}_t A^{(t)}(z)}{A^{(t)}(z)} = f(z) \end{aligned}$$

where we have used (6) and (21) in (a), and (23) in (b).  $\square$

*Lemma 5:* The inverse function  $f^{-1}$  fulfills

$$f^{-1}(M - \alpha) = \frac{1}{f^{-1}(\alpha)} \quad (24)$$

$\forall \alpha \in (0, 1)$  if and only if  $A^{(t)}(z)$  is symmetric for every  $t \in I_c$ .

*Proof:* By Lemma 4, the function  $f$  fulfills (22) if and only if  $A^{(t)}(z)$  is symmetric for every  $t \in I_c$ . By applying the inverse function to both sides of (22) and by letting  $f(z^{-1}) = \alpha$  for all  $z \in \mathbb{R}^+ \setminus \{0\}$ , we obtain the statement.  $\square$

*Theorem 2 (Sufficient condition for symmetry):* Consider a GLDPC code ensemble with a regular VN set, composed of repetition codes all of length  $q$ , and a hybrid CN set, composed of a mixture of  $n_c$  different linear block codes. If  $A^{(t)}(z)$  is symmetric for each  $t \in I_c$ , then the spectral shape of the ensemble fulfills

$$G(M - \alpha) = G(\alpha) \quad (25)$$

for all  $\alpha \in (0, 1)$ .

*Proof:* By Lemma 2, the hypothesis that  $A^{(t)}(z)$  is symmetric for each  $t \in I_c$  implies  $\bar{u}_t = s_t$  for every  $t \in I_c$ , and therefore  $M = 1$ . From (8) we have:

$$\begin{aligned} G(M - \alpha) &= (1 - q)h(M - \alpha) - q(M - \alpha) \log f^{-1}(M - \alpha) \\ &\quad + q \left( \int \rho \right) \sum_{t \in I_c} \gamma_t \log A^{(t)}(f^{-1}(M - \alpha)) \\ &\stackrel{(a)}{=} (1 - q)h(M - \alpha) - q(M - \alpha) \log \frac{1}{f^{-1}(\alpha)} \\ &\quad + q \left( \int \rho \right) \sum_{t \in I_c} \gamma_t \log A^{(t)} \left( \frac{1}{f^{-1}(\alpha)} \right) \\ &\stackrel{(b)}{=} (1 - q)h(M - \alpha) - q\alpha \log(f^{-1}(\alpha)) \\ &\quad + q \left( \int \rho \right) \sum_{t \in I_c} \gamma_t \log A^{(t)}(f^{-1}(\alpha)) \\ &= G(\alpha) \end{aligned}$$

where (a) follows from symmetry of  $A^{(t)}(z)$  and Lemma 5, (b) from symmetry of  $A^{(t)}(z)$ , Lemma 3 and (6), and the final line from  $M = 1$ .  $\square$

*Theorem 3 (Necessary condition for symmetry):* Consider a GLDPC code ensemble with a regular VN set, composed of repetition codes all of length  $q$ , and a hybrid CN set, composed of a mixture of  $n_c$  different linear block codes. If the spectral shape fulfills  $G(M - \alpha) = G(\alpha)$  for all  $\alpha \in (0, 1)$ , then  $M$  is a fixed point of the function

$$\Gamma(x) = 2 f \left( \left( \frac{x}{2 - x} \right)^{\frac{q-1}{q}} \right). \quad (26)$$

*Proof:* The proof is somewhat lengthy and we only sketch it due to space constraints. The first step consists of developing an expression for  $F(\alpha) = G(M - \alpha) - G(\alpha)$ . The obtained expression can be written as a function of  $z$ , denoted by  $J(z)$ , by defining  $z = f^{-1}(\alpha)$ . We must have  $J(z) = 0$  for all  $z \in (0, +\infty)$ , and therefore  $J'(z) = 0$  for all  $z \in (0, +\infty)$ . This latter condition must hold in particular for  $z = f^{-1}(\frac{M}{2})$ , in which case we obtain the compact condition

$$q \log f^{-1} \left( \frac{M}{2} \right) = (1 - q) \log \left( \frac{2 - M}{M} \right)$$

which is equivalent to

$$M = 2 f \left( \left( \frac{M}{2 - M} \right)^{\frac{q-1}{q}} \right).$$

$\square$

Note that if  $A^{(t)}(z)$  is symmetric for each  $t \in I_c$  then we have  $M = 1$ , which is always a fixed point of  $\Gamma(x)$  defined in (26) since it is possible to show that  $f(1) = \frac{1}{2}$ .

## V. EXAMPLES

*Example 1 (Tanner code with (7, 4) Hamming CNs):*

Consider a rate  $R = 1/7$  Tanner code ensemble where all VNs have degree 2 and where all CNs are (7, 4) Hamming codes (it was shown in [14], [15] that this ensemble has good spectral shape behavior). The WEF of a Hamming (7, 4) CN is given by  $A(z) = 1 + 7z^3 + 7z^4 + z^7$ , while its local MAP-SSEF and



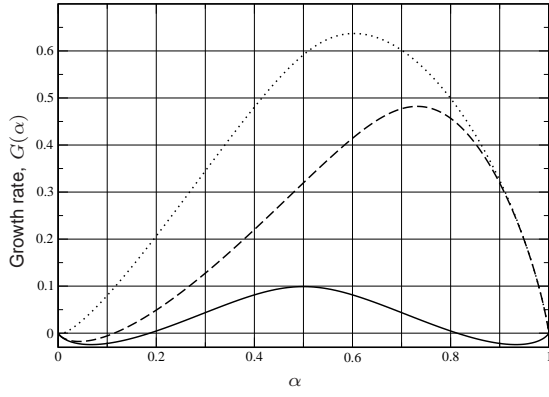


Fig. 1. Spectral shapes of the Tanner code ensemble in Example 1. Solid: weight spectral shape (relative minimum distance:  $\alpha^* = 0.18650$ ). Dashed: stopping set size spectral shape under MAP decoding at the CNs (relative minimum stopping set size:  $\alpha_\Phi^* = 0.11414$ ). Dotted: stopping set size spectral shape under BD decoding at the CNs (relative minimum stopping set size:  $\alpha_\Psi^* = 0.01025$ ).

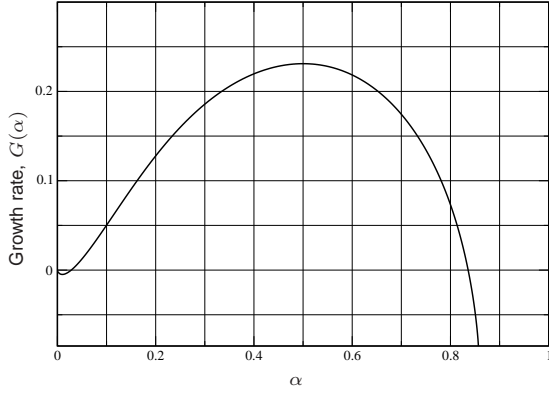


Fig. 2. Weight spectral shape of the check-hybrid GLDPC code ensemble in Example 2. Relative minimum distance:  $\alpha^* = 0.028179$ .

BD-SSEF are given by  $\Phi(z) = 1 + 7z^3 + 10z^4 + 21z^5 + 7z^6 + z^7$  and  $\Psi(z) = 1 + 35z^3 + 35z^4 + 21z^5 + 7z^6 + z^7$  respectively. Note that we have  $M = \frac{q}{s} = 1$  in all three cases. A plot of  $G(\alpha)$ ,  $G_\Phi(\alpha)$  and  $G_\Psi(\alpha)$  obtained by implementation of (18) is depicted in Fig. 1. We observe that  $A(z)$  satisfies the conditions of Theorem 2. This is reflected by the fact that the weight spectral shape  $G(\alpha)$  is symmetric with respect to  $\alpha = 1/2$ .

*Example 2 (Check-hybrid ensemble):* Consider a rate  $R = 1/3$  check-hybrid GLDPC code ensemble where all VNs are repetition codes of length  $q = 3$  and whose CN set is composed of a mixture of two linear block code types ( $I_c = \{1, 2\}$ ). CNs of type 1  $\in I_c$  are length-7 SPC codes with WEF  $A^{(1)}(z) = [(1+z)^7 + (1-z)^7]/2$  and  $\gamma_1 = 0.722$ , while CNs of type 2  $\in I_c$  are  $(7, 4)$  codes with WEF  $A^{(2)}(z) = 1 + 5z^2 + 7z^4 + 3z^6$  and  $\gamma_2 = 0.278$ . The weight spectral shape of this ensemble, obtained from (8), is depicted in Fig. 2. Note that for this ensemble,  $M = 6/7$ . This value is not a fixed point of the function  $F(x)$  defined in (26) (the only fixed point between 0 and 1 is  $x = 0.888421$ ). As expected, the weight spectral shape does not exhibit any symmetry property.

*Example 3 (Ensemble with bad spectral shape behavior):*

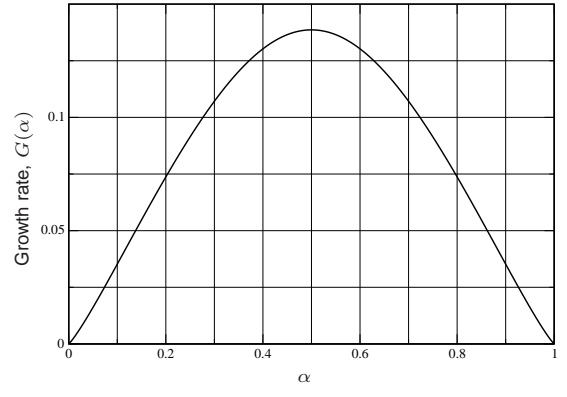


Fig. 3. Weight spectral shape of the Tanner code ensemble in Example 3.

Consider a rate  $R = 1/5$  Tanner code ensemble where all VNs are repetition codes of length  $q = 2$  and where all CNs are  $(5, 3)$  linear block codes with WEF  $A(z) = 1 + 3z^2 + 3z^3 + z^5$ . This ensemble is known to have bad spectral shape behavior ( $\alpha^* = 0$ ) since we have  $\lambda'(0)C = 6/5 > 1$ , where  $\lambda(x) = x$  and  $C = 2A_2/s$  [12], [13]. A plot of the weight spectrum for this ensemble, obtained from (18) is depicted in Fig. 3. We observe that the plot of  $G(\alpha)$  is symmetric, due to the fact that  $A(z)$  is symmetric ( $M = 1$ ). As expected, the derivative of  $G(\alpha)$  at  $\alpha = 0$  is positive and hence  $\alpha^* = 0$ .

## VI. CONCLUSION

A simple expression has been developed for both the weight and the stopping set size spectral shape of GLDPC code ensembles with a regular VN set and a hybrid CN set. Some known results (specifically, an expression for the spectral shape of regular LDPC codes and an asymptotic expression of the spectral shape for GLDPC codes as the normalized weight tends to zero) follow as corollaries of the developed formula. Moreover, symmetry properties of the spectral shape function have been discussed. A sufficient condition and a necessary condition for the weight spectral shape function to be symmetric have been identified.

## APPENDIX I

### ASYMPTOTIC CASE $\alpha \rightarrow 0$

For small  $\alpha$ , the expression (8) reduces to a known formula first developed in [12] for Tanner codes, and extended in [13] to irregular GLDPC codes. This formula is here obtained as a simple corollary of Theorem 1.

*Corollary 2:* In the limit where  $\alpha \rightarrow 0$ , the growth rate of the weight distribution of a GLDPC code ensemble with a regular VN set fulfills

$$G(\alpha) \rightarrow \left(q - \frac{q}{r} - 1\right)\alpha \log \alpha + \frac{q\alpha}{r} \log \left(er \int \rho \sum_{t \in X_c} \gamma_t A_r^{(t)}\right). \quad (27)$$

where  $r$  denotes the smallest minimum distance of all CN types (i.e.  $r \triangleq \min_{t \in I_c} r_t$ ),  $X_c$  denotes the set of CN types

with this minimum distance (i.e.  $X_c = \{t \in I_c : r_t = r\}$ ) and  $e$  denotes Napier's number.

*Proof:* Let  $\alpha = f(z_0)$ . From the definition of  $f$  given in (7), it is readily shown that if  $\alpha \rightarrow 0$ , we must have  $z_0 \rightarrow 0$ . Next, note that in (7), each term  $t \in I_c$  in the sum on the right-hand side is a rational polynomial in  $z_0$  whose denominator tends to unity as  $\alpha \rightarrow 0$ , and whose numerator is dominated as  $\alpha \rightarrow 0$  by the term corresponding to the lowest power of  $z_0$ . Therefore, as  $\alpha \rightarrow 0$ , (7) becomes  $\alpha \rightarrow (\int \rho) \sum_{t \in X_c} \gamma_t r A_r^r z_0^r$ , or equivalently,

$$z_0 \rightarrow \left( \frac{\alpha}{r \int \rho \sum_{t \in X_c} \gamma_t A_r^{(t)}} \right)^{1/r}. \quad (28)$$

Using  $\log(1+x) \rightarrow x$  as  $x \rightarrow 0$ , we have that for every  $t \in I_c$   $\log A^{(t)}(z_0) \rightarrow A_{r_t}^{(t)} z_0^{r_t}$  as  $\alpha \rightarrow 0$ , and so

$$\sum_{t \in I_c} \gamma_t \log A^{(t)}(z_0) \rightarrow \sum_{t \in X_c} \gamma_t A_r^{(t)} z_0^r = \frac{\alpha}{r \int \rho}. \quad (29)$$

Note that in these steps we have again used the fact that a polynomial expression in  $z_0$  is dominated by its lowest degree term as  $z_0 \rightarrow 0$ . Next, observe that  $h(\alpha) \rightarrow -\alpha \log \alpha$  as  $\alpha \rightarrow 0$  and therefore we obtain (using (28) and (29) in (8))

$$G(\alpha) \rightarrow (q-1) \alpha \log \alpha - \frac{q \alpha}{r} \log \left( \frac{\alpha}{r \int \rho \sum_{t \in X_c} \gamma_t A_r^{(t)}} \right) + \left( q \int \rho \right) \left( \frac{\alpha}{r \int \rho} \right)$$

which coincides with (27).  $\square$

## APPENDIX II

### CLOSED FORM EXPRESSIONS FOR THE GROWTH RATE

It is worthwhile to note that in some cases, (18) can be expressed in closed form because  $f^{-1}(\alpha)$  can be expressed analytically. This is the case, for instance, for the (3,6) regular LDPC ensemble, for which  $f(z) = \alpha$  becomes  $ax^3 + bx^2 + cx + d = 0$ , where  $x = z^2$  and  $(a, b, c, d) = (\alpha - 1, 15\alpha - 10, 15\alpha - 5, \alpha)$ . This cubic equation in  $x$  may be solved by Cardano's method (see, e.g., [17, p. 17]; the discriminant  $\Delta = \rho^3 + \mu^2$  is negative for every  $\alpha \in (0, 1)$ , where

$$\rho = \frac{3ac - b^2}{9a^2}; \quad \mu = \frac{9abc - 27a^2d - 2b^3}{54a^3}.$$

The required solution is then uniquely and analytically identified as  $f^{-1}(\alpha) = z = \sqrt{x}$  where  $x = 2\sqrt{-\rho} \cos(\theta/3) - \frac{b}{3a} > 0$  and  $\theta = \tan^{-1}(\sqrt{-\Delta}/\mu)$ .

Similarly, the weight spectral shape of a (4,8) regular LDPC ensemble may be expressed in closed form through the solution of a quartic equation.

## REFERENCES

- [1] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inf. Theory*, vol. 27, no. 5, pp. 533–547, Sept. 1981.
- [2] R. G. Gallager, *Low-density parity-check codes*. Cambridge, Massachusetts: M.I.T. Press, 1963.
- [3] G. Liva, W. Ryan, and M. Chiani, "Quasi-cyclic generalized LDPC codes with low error floors," *IEEE Trans. Commun.*, vol. 56, no. 1, pp. 49–57, Jan. 2008.
- [4] Y. Wang and M. Fossorier, "Doubly generalized LDPC codes," in *Proc. of IEEE 2006 Int. Symp. on Information Theory*, Seattle, WA, USA, July 2006, pp. 669–673.
- [5] M. F. Flanagan, E. Paolini, M. Chiani and M. P. C. Fossorier, "Growth rate of the weight distribution of doubly-generalized codes: General case and efficient evaluation," in *Proc. of 2009 IEEE Global Telecommunications Conf.*, Honolulu, HI, USA, Nov./Dec. 2009.
- [6] S. Abu-Surra, W. E. Ryan and D. Divsalar, "Ensemble enumerators for protograph-based generalized LDPC codes," in *Proc. of 2007 IEEE Global Communications Conf.*, Washington, DC, USA, Nov. 2007, pp. 1492–1497.
- [7] K. Kasai, C. Poulliat, D. Declercq, T. Shibuya and K. Sakaniwa, "Weight distribution of non-binary LDPC codes," in *Proc. of 2008 Int. Symp. on Inf. Theory and its Applications*, Auckland, New Zealand, Dec. 2008, pp. 748–753.
- [8] K. Kasai, T. Awano, D. Declercq, C. Poulliat and K. Sakaniwa, "Weight distributions of multi-edge type LDPC codes," in *Proc. of 2009 IEEE Int. Symp. on Inf. Theory*, Seoul, Korea, June/July 2009, pp. 60–64.
- [9] M. Lentmaier, M. B. S. Tavares, and G. P. Fettweis, "Exact erasure channel density evolution for protograph-based generalized LDPC codes," in *Proc. of 2009 IEEE Int. Symp. on Inf. Theory*, Seoul, Korea, June/July 2009, pp. 566–570.
- [10] C. -Li Wang, S. Lin and M. P. C. Fossorier, "On asymptotic ensemble weight enumerators of multi-edge type codes," in *Proc. of 2009 IEEE Global Telecommunications Conf.*, Honolulu, HI, USA, Nov./Dec. 2009.
- [11] C. Di, D. Proietti, I. E. Telatar, T. J. Richardson and R. Urbanke, "Finite-length analysis of low-density parity-check codes on the binary erasure channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1570–1579, June 2002.
- [12] J. -P. Tillich, "The average weight distribution of Tanner code ensembles and a way to modify them to improve their weight distribution," in *Proc. of 2004 IEEE Int. Symp. on Information Theory*, Chicago, IL, USA, June/July 2004, pp. 7.
- [13] E. Paolini, M. Chiani and M. Fossorier, "On the growth rate of GLDPC codes weight distribution," in *Proc. of 2008 IEEE Int. Symp. on Spread Spectrum Techniques and Applications*, Bologna, Italy, Aug. 2008, pp. 790–794.
- [14] M. Lentmaier and K. Zigangirov, "On generalized low-density parity-check codes based on Hamming component codes," *IEEE Commun. Lett.*, vol. 3, no. 8, pp. 248–250, Aug. 1999.
- [15] J. J. Boutros, O. Pothier and G. Zémor, "Generalized low density (Tanner) codes," in *Proc. of 1999 IEEE Int. Conf. on Communications*, vol. 1, Vancouver, Canada, June 1999, pp. 441–445.
- [16] A. Orlitsky, K. Viswanathan, and J. Zhang, "Stopping set distribution of LDPC code ensembles," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 929–953, Mar. 2005.
- [17] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. New York: Dover, 1964.